

# Digitizing Standards with OsteoSurvey:

## A case study in open access data collection at Deir el-Medina, Egypt

Anne Austin, Mellon Postdoctoral Fellow at Stanford University

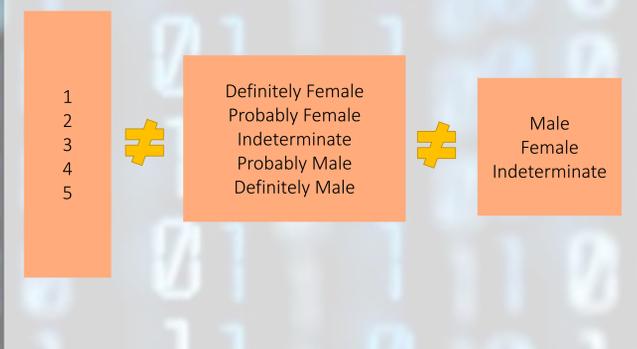
### INTRODUCTION

In the early 1990s, bioarchaeology as a discipline encountered two rising issues: (1) the repatriation of human remains and (2) incompatible datasets between research projects. Using basic recommendations to address these problems (Rose et al., 1991), *Standards for Data Collection from Human Skeletal Remains* (Buikstra and Ubelaker, 1994) allowed researchers to quickly record and compare their data. Over 20 years later, *Standards* is still an essential tool for bioarchaeologists across the world as well as a pillar of the bioarchaeological training employed at the ASU Bioarchaeology Field School in Kampsville, IL. Yet, despite the widespread use of *Standards* and numerous methodological additions, we continue to face major obstacles when reusing bioarchaeological data. This poster elucidates existing problems with data reuse in bioarchaeology and offers best practices for data collection and publication to enable you and future researchers to access and reuse pre-existing datasets.

### PROBLEMS IN BIOARCHAEOLOGICAL DATA REUSE

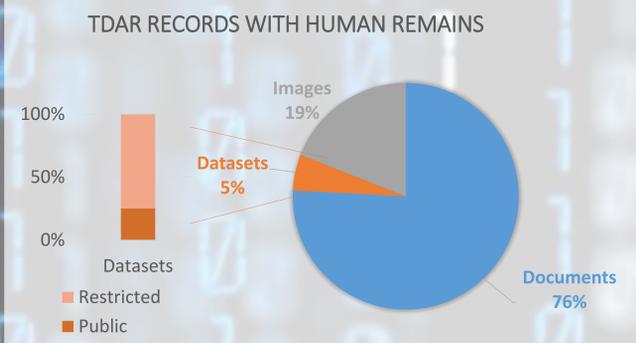
#### 1.) Incompatible data schemas

Even basic categories, such as sex and age, cannot always be mapped directly between most datasets due to great amounts of variation in encoding terminology:



#### 2.) A lack of access to original datasets

Printed publication formats make it difficult to publish complete datasets and many scholars are reticent to publish their data online prior to completing ALL potential analyses. Currently, only 1% of bioarchaeological files stored at tDAR are publically-accessible datasets.



#### 3.) Observations vs. Interpretations

Publications often include interpretations of the remains (e.g., age category), but not initial observations (e.g., epiphyseal fusion). When reusing a dataset, we require these initial observations to reinterpret the material. This zooarchaeological example from Atici and colleagues (2013) shows how much variation results in data reuse when published datasets only include interpreted age and not initial observations of epiphyseal fusion:

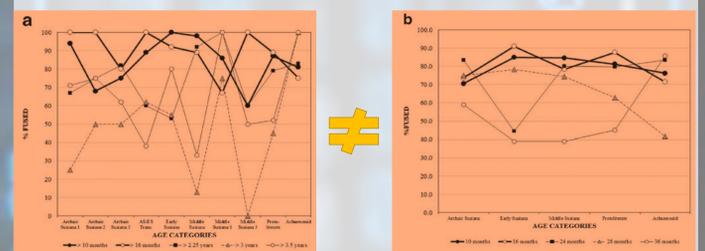


Figure 2a-b from Atici et al. (2013)

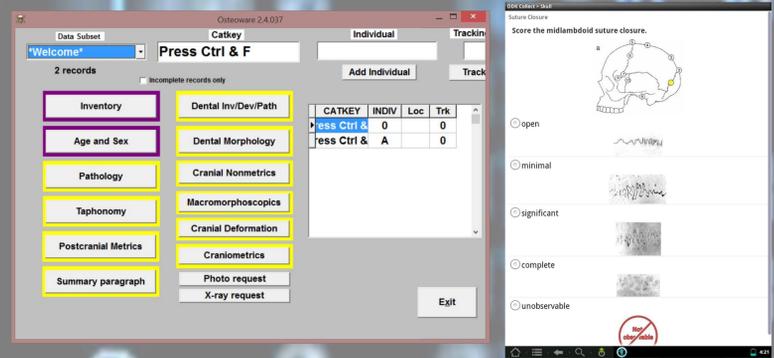
### WE ARE THE 99%: HOW TO INCREASE DATA REUSE IN BIOARCHAEOLOGY?

The above issues in data reuse can only be resolved if (a) more bioarchaeologists choose to publish their datasets and (b) we standardize data publication to the same extent as we do data collection. Increased data reuse benefits researchers because:

1. Dataset publication satisfies the data management legal mandate of funding agencies like NSF
2. Online datasets can lead to increased publications through future collaborations
3. Data reuse enables us to answer broader diachronic and interregional research questions (e.g., Arbuckle et al. 2014)
4. Digital online repositories permanently back-up your data

Online databases such as the Wellcome Osteological Research Database (WORD) have begun to address these issues (Giesen 2013), but how can individual research projects create and publish digitized datasets? Best practices begin at collection: digital data collection tools, like Osteoware or OsteoSurvey, enable individual researchers to address these issues from the moment their data are born. Using my work at Deir el-Medina with *OsteoSurvey*, I illustrate how digital data collection tools can avoid the above problems and streamline dataset publication.

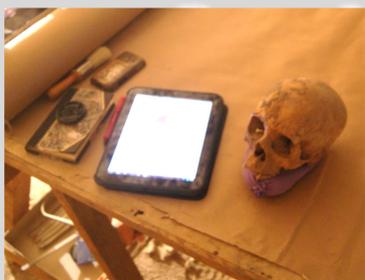
*Osteoware* (left) and *OsteoSurvey* (right) used for data collection



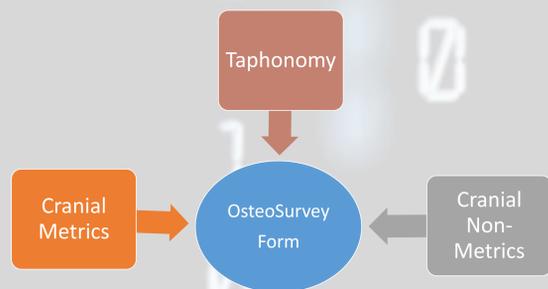
### BEST PRACTICES IN DATA COLLECTION: A CASE STUDY AT DEIR EL-MEDINA USING OSTEOSURVEY

#### A FIELD-TESTED DIGITAL DATA COLLECTION MODEL

Deir el-Medina is the village of the workmen who cut and decorated the Egyptian royal tombs in Egypt during the New Kingdom (1550-1070 B.C.E.). Bioarchaeological data collection at Deir el-Medina requires working inside the tombs in order to inventory and document commingled remains. To accommodate field-based data collection, I developed *OsteoSurvey*: a free set of forms based on *Standards* that works with Open Data Kit (ODK) on all Android devices. In order to manage data reuse, *OsteoSurvey* employs modular formatting, skip logic, and data validation to increase both the quantity and quality of data collected, while also enabling customized data management.



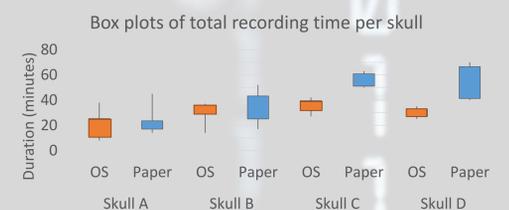
*OsteoSurvey* can be employed in the field, ensuring all observations are 'born' digital.



Modular data organization allows researchers to select which methods to use, add new methods, and publish select modules.



Digitizing all observations enables one to publish a raw dataset, rather than just interpretations of the data.



Transcription Errors			
Survey type	Present	Absent	Total
Paper forms	25	7	32
OsteoSurvey	4	28	32

Skip logic increases recording speed, while data validation ensures fewer accidentally missed fields (Austin 2014).

#### 1.) Field-based Data Collection

#### 2.) Modular Formatting

#### 3.) Digitized Observations

#### 4.) Automatic Data Entry Constraints

### BEST PRACTICES IN ONLINE PUBLICATION

When employing *OsteoSurvey* at Deir el-Medina, I also streamlined dataset publication by exporting data in a durable .CSV format, incorporating metadata into dataset publication, and using a short-term embargo for dataset publication on tDAR to ensure future access.

Average lifespan is less than 5 years for physical storage (Rothenberg 1995).



*OsteoSurvey* exports data as .CSV



Physical media and file types face short-term obsolescence. Storage in digital repositories (e.g., tDAR) using simple file types like .CSV ensure long-term access for future reuse.

#### 1.) Durable File Types and Storage



*OsteoSurvey* includes metadata files based on *Standards* and *Dublin Core*

Co-publishing metadata with datasets enables future researchers to decode datasets.

#### 2.) Accompanying Metadata

Embargo options at tDAR



Module	Dataset Publication
Taphonomy	Public
Cranial Metrics	2 year embargo
Cranial Non-metrics	Restricted

Embargos allow us to publish peer-reviewed journal publications prior to making datasets available for reuse. *OsteoSurvey* also publishes to a secured server so that multiple researchers can contribute to and view a group dataset online.

#### 3.) Publishing Options

### TRY OSTEOSURVEY

For more information, go to:  
<http://www.anneaustin.com/osteosurvey/>



### Works Cited

Arbuckle BS, Kansa SW, Kansa EC, Orton D, Çakırlar C, Gourichon L, Atici L, Galik A, Marciniak A, Mulville J, Buitenhuis H, Carruthers D, De Cupere B, Demiregi A, Frane S, Helmer D, Martin L, Peters J, Pöllath N, Pawłowska K, Russell N, Twiss K, Würtenberger D. 2014. Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey. *PLoS ONE* 9 (9):e99845.

Atici L, Kansa SW, Lev-Tov J, Kansa EC. 2013. Other People's Data: A Demonstration of the Imperative of Publishing Primary Data. *Journal of Archaeological Method and Theory* 20 (4): 663-681.

Austin A. 2014. Mobilizing Archaeologists: Increasing the Quantity and Quality of Data Collected in the Field with Mobile Technology. *Advances in Archaeological Practice: A Journal of the Society for American Archaeology* 2: 13-23.

Buikstra JE, Ubelaker DH. 1994. *Standards for Data Collection from Human Skeletal Remains*. Fayetteville: Arkansas Archeological Survey No 44.

Giesen M, McCarrison K, Park V. 2013. Dead and Forgotten? Some Observations on Human Remains Documentation in the UK. In: Giesen M, editor. *Curating Human Remains: Caring for the Dead in the United Kingdom*. Suffolk: Boydell Press. p. 53-64.

Rose JC, Anton S, Aufderheide AC, Buikstra JE, Eisenberg LE, Gregg JB, Neiburger EJ, Rothschild BM. 1991. *Paleopathology Association Skeletal Database Committee recommendations*. Detroit: Paleopathology Association.

Rothenberg J. 1995. Ensuring the longevity of digital documents. *Scientific American* 272: 42-47.